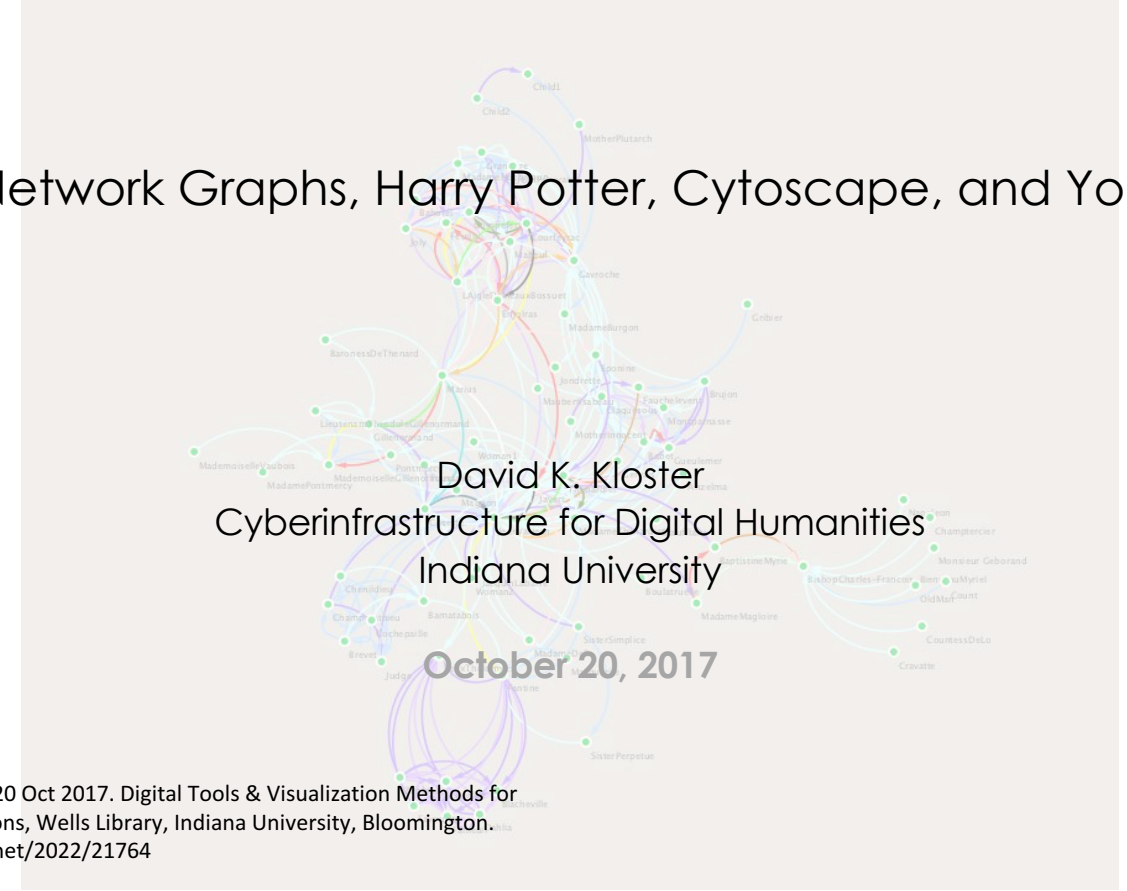
Network Graphs, Harry Potter, Cytoscape, and You!



David K. Kloster
Cyberinfrastructure for Digital Humanities
Indiana University

October 20, 2017

Kloster, David. "Network Graphs." 20 Oct 2017. Digital Tools & Visualization Methods for Humanists Series. Scholars' Commons, Wells Library, Indiana University, Bloomington. Retrieved from: http://hdl.handle.net/2022/21764

Download Cytoscape

www.cytoscape.org

Follow Along

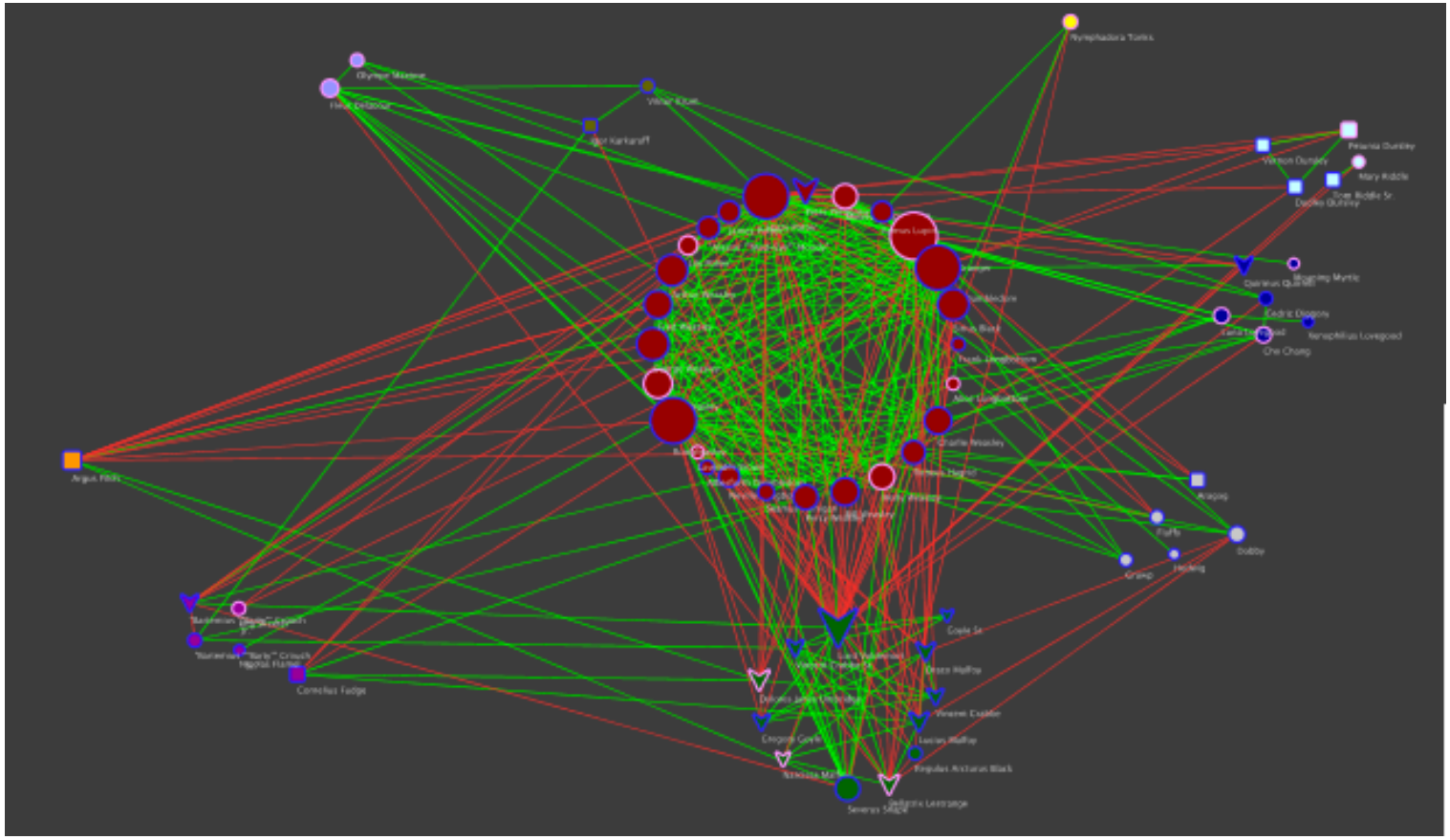
https://iu.box.com/v/network-graphs2017Fa

Digital Arts and Humanities Workshop Series – Fall 2017

Fridays @ noon -- Scholars Commons IQ-Wall

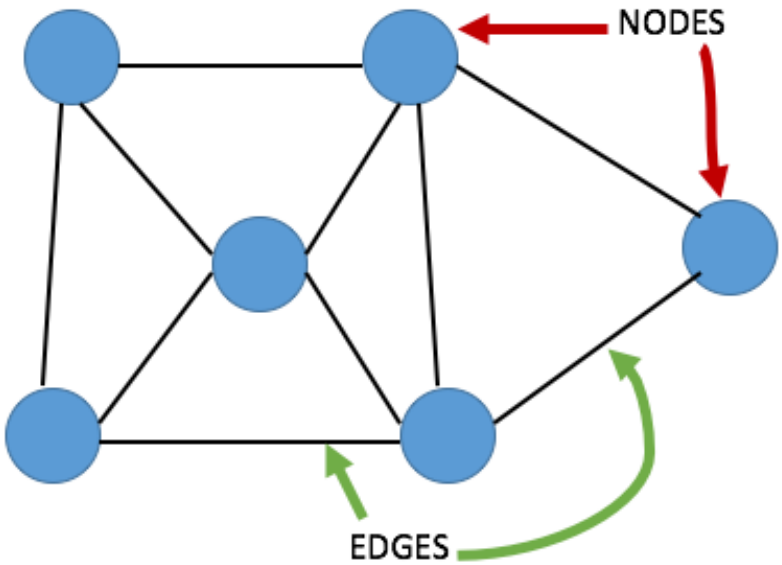
Date	Topic	Presenter
Aug. 25	Intro to Visualization	Michael Boyles
Sep. 1	Intro to Digital Humanities	Tassie Gniady
Sep. 8	Virtual Reality	Bill Sherman
Sep. 15	Intro to R	Tassie Gniady
Sep. 22	Advanced Media	Chris Eller
Sep. 29	Augmented Reality	Chauncey Frend
Oct. 13	R for Text	Tassie Gniady
Oct. 20	Network Graphs	David Kloster
Oct. 27	R for Twitter	Tassie Gniady
Nov. 3	3D Scanning and Printing	Jeff Rogers
Nov. 10	3D Photogrammetry	Tassie Gniady
Dec. 1	IQ-Tables & Touch-Enabled Software Workflows	David Reagan

What is a Network Graph and Why Make One?

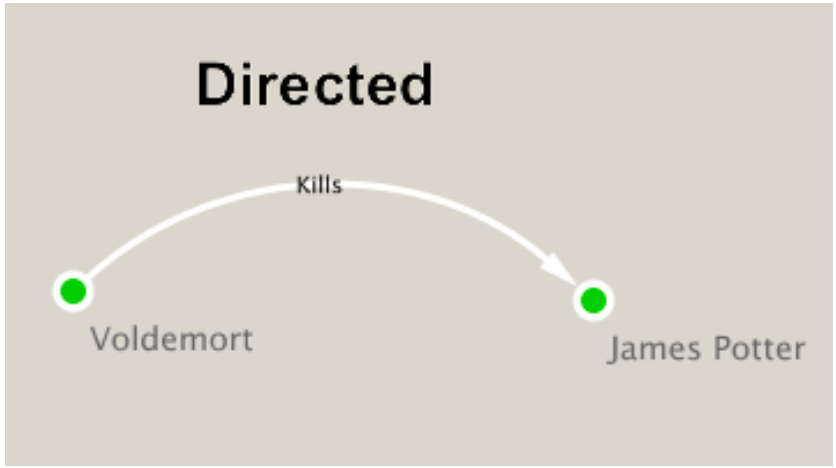


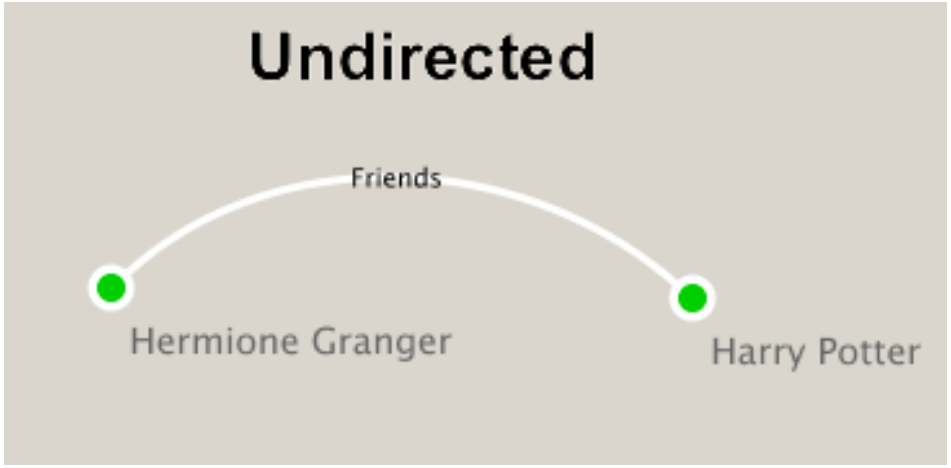
Network Graphs…The Basics

- Nodes (vertices) = Objects
- Edges = The relationship(s)

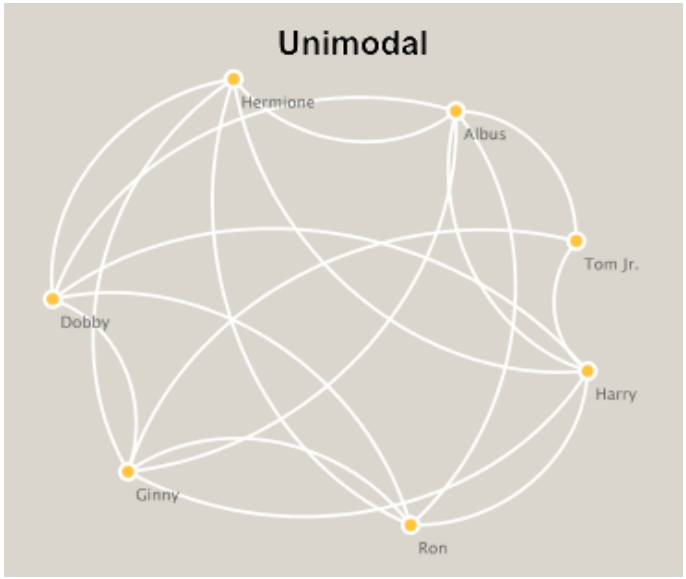


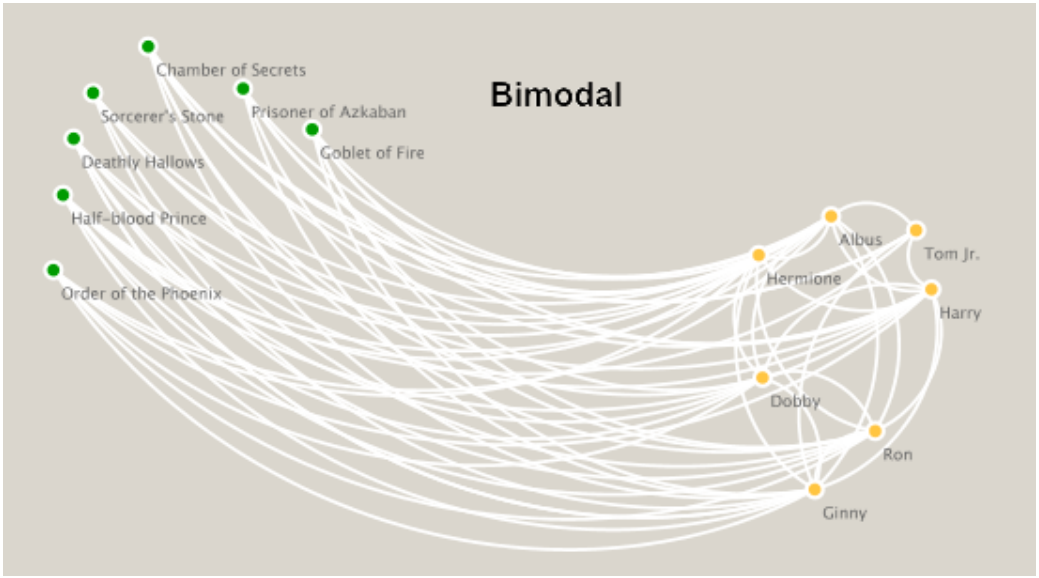
Terms to Know



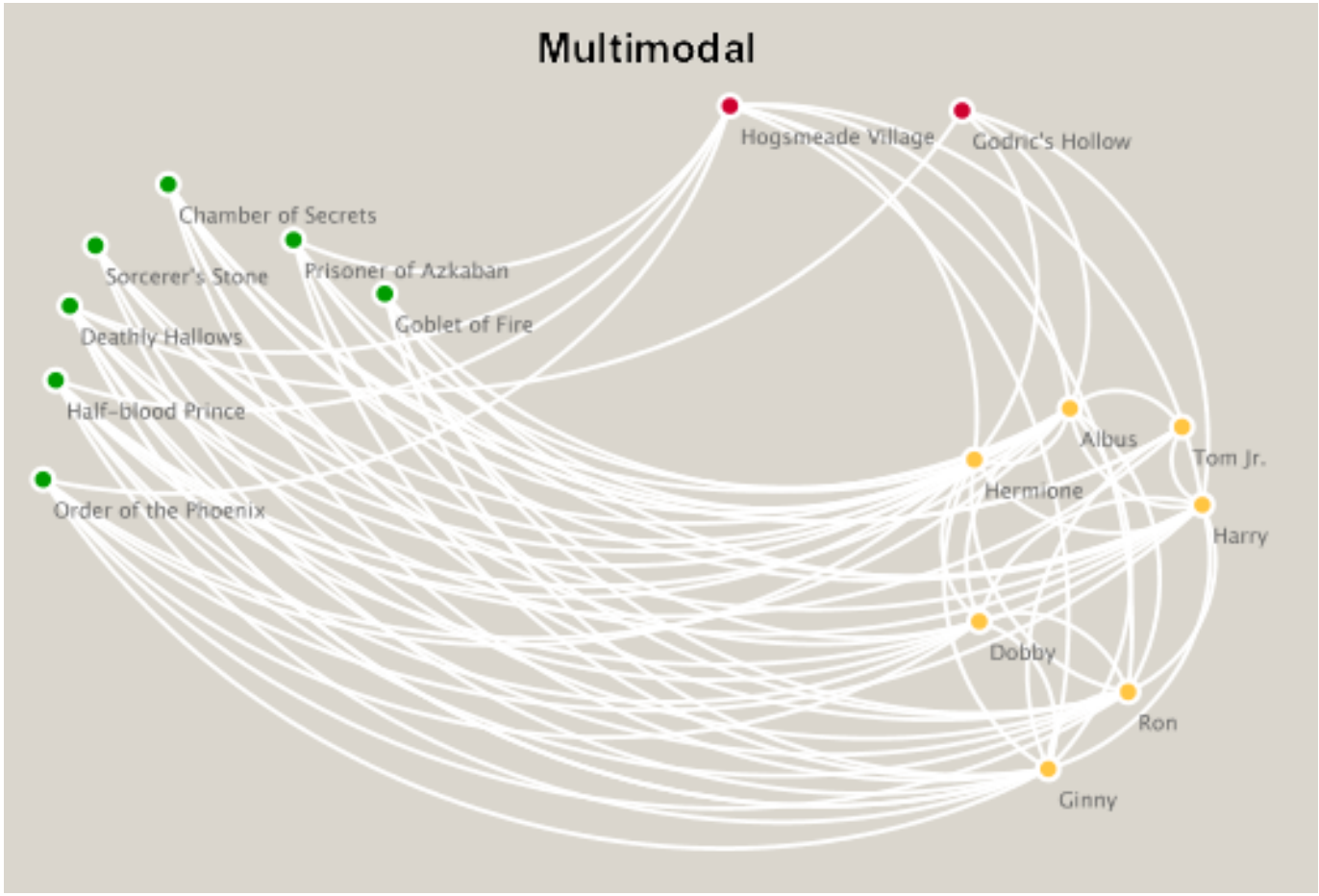


Types of Networks





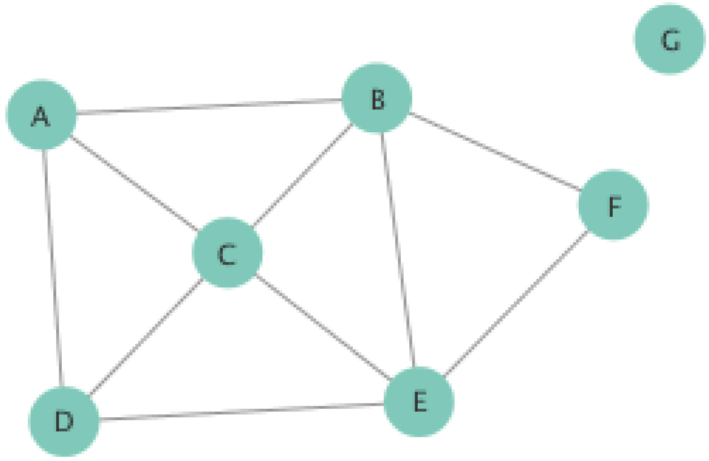
Types of Networks



Node Attributes

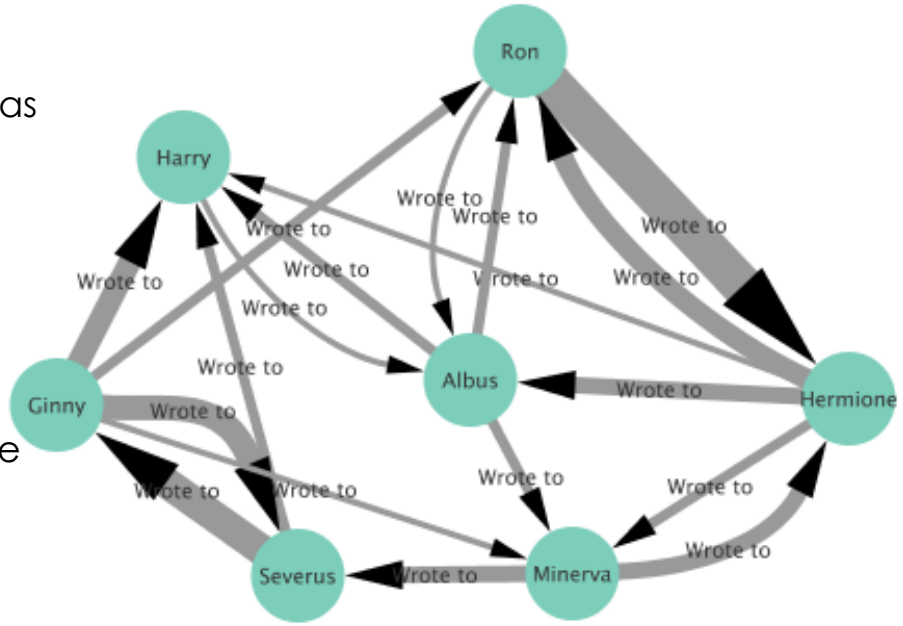
- Degree - number of edges that connect to a node. For example, **Node C** has a degree of four and **Node F** only has a degree of 2.

- Isolates – Nodes not connected to any others edges. **Node G** is an isolate.



Attributes Cont…

- In-degree - is calculated by determining the number of edges that point to a node
  - **Ron** has an in-degree of 3 and **Ginny** only has an in-degree of 1.
- Out-degree – is calculated by determining the number of edges that point away from a node
  - **Hermione** has an out-degree of 4 and **Harry** has an out-degree of 1.
- Weight (edge attribute) – Strength of the tie by the thickness of the edges between nodes
  - In the example, the edge between **Ron** and **Hermione** is Strongest (Wrote most letters).



Popular Network Software

SOFTWARE	Ease of Use	Pros	Cons
Cytoscape	• No programming knowledge required • Minimal understanding of networks • Basic excel knowledge	• Scales with data • User friendly • Lots of options • Well supported	• Made for hard sciences
igraph	• Requires use of R, Python, or C	• Can be used with multiple programming languages • Allows for more manipulation of network aesthetics	• Requires some programming knowledge (we can help with that)
Gephi	• Intermediate knowledge of excel needed (ex. VLOOKUP)	• Allows for aesthetically pleasing networks • Intuitive GUI	• Not supported well (4 years between updates) • Still has reported issues with Macs
Palladio	• Easy to follow tutorials on the web page • No programming knowledge required	• More options • User friendly instructions	• Mostly intended for geospatial networks
Fusion Tables	• Uses .csv, .tsv, .txt, .kml or google sheets	• Basic • Easy to use	• Minimal options

Info and Examples on Networks in the Humanities

Scott Weingart

[Demystifying Networks Parts I & II. Journal of Digital Humanities.]

Miriam Posner

- [Cytoscape Tutorial]
- [Blog post about Cytoscape]

Rachael Cohen & Angie Thorpe

[IU Bloomington and Kokomo library Search terms by subject]

Download Cytoscape

www.cytoscape.org

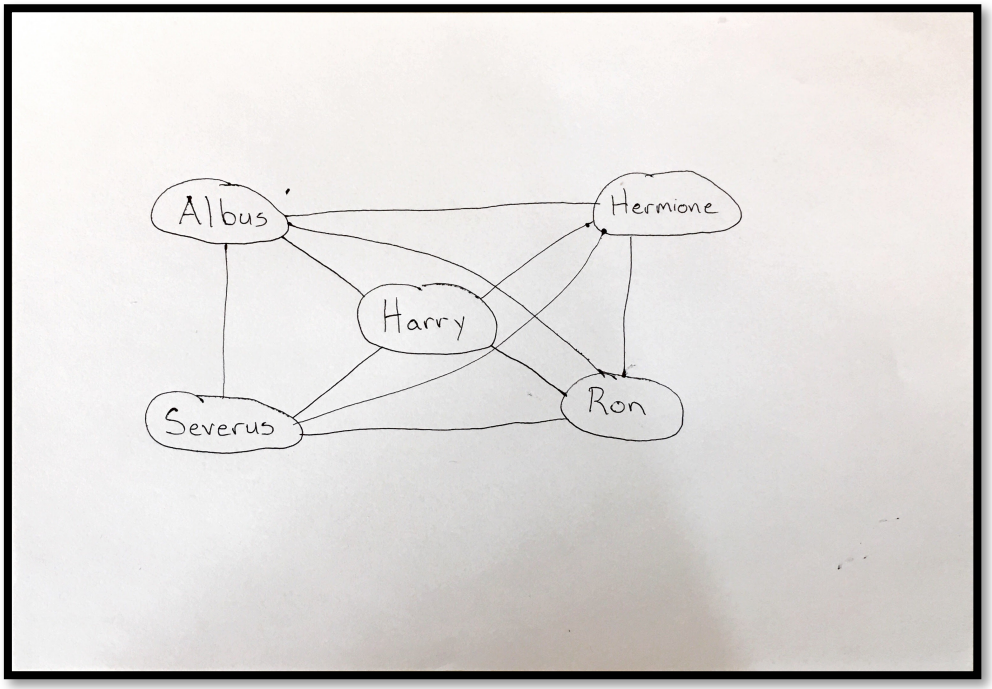
Creating a Network
(General Steps)

- Determine your problem/scholarly question
- Assess what data you have and what is available
- Make analog network graph on paper
- Organize and clean your data (determine what your objects are and the nature of their relationship)
- Plug your data into your software (in our case Cytoscape)
- Map your network graph (assign node colors, node size, edge weights, etc…)
- Analyze and decide if the network solves your problem/answers your question

NOTE: Should practice on something you know really well first. (So you can see if your results are making sense)

Network Creation using Cytoscape

- Steps 1 & 2 already done
- Make analog network graph



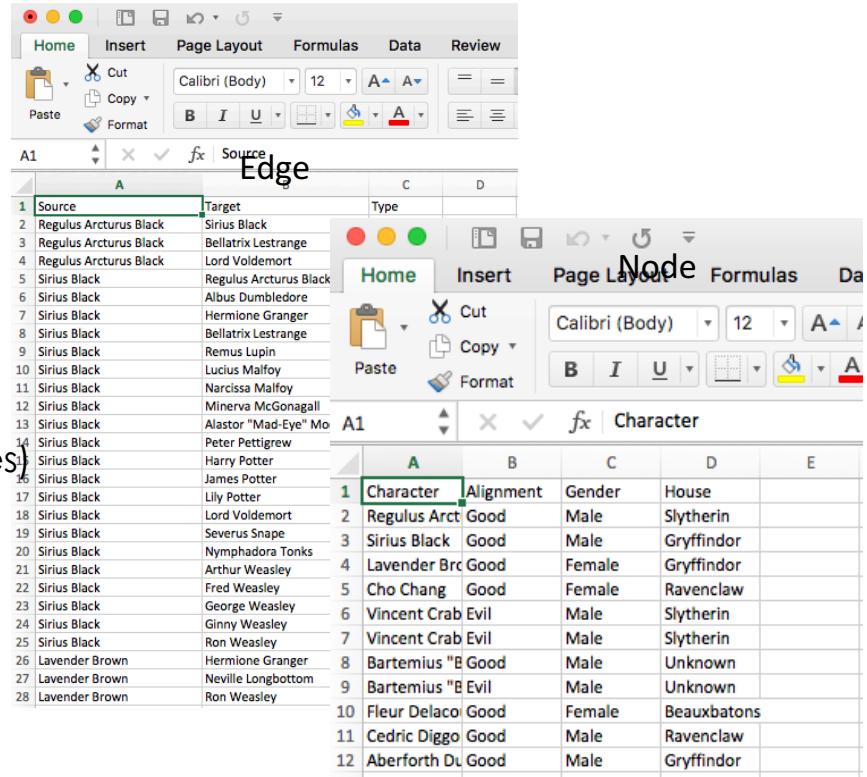
Edge and Node List

- Organize
  - In a spreadsheet:
    - Create an edge list (determines objects, their relationship to one another, and edge attributes)

    - Create a node list (determines object attributes)

    (**NOTE:** Make sure spelling and capitalization is exact and universal in both lists)

    Normally, this would be the next step, however…



Download Google Sheets Sample

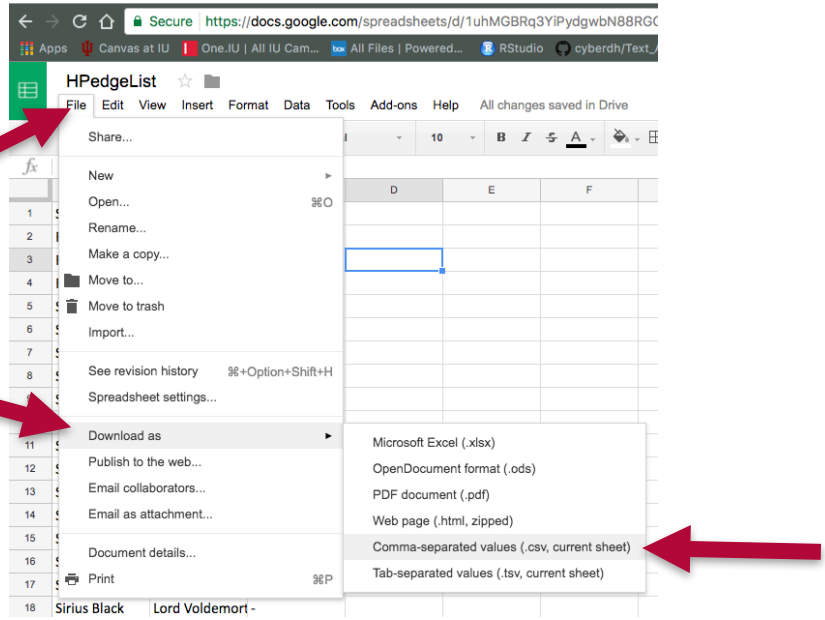
To save time, download this edge list here:
http://bit.ly/edgelisthp

And this node list here:
http://bit.ly/hpnodelist

Go to File>Download as>Comma-separated values (.csv, current sheet) for both the edge and node list

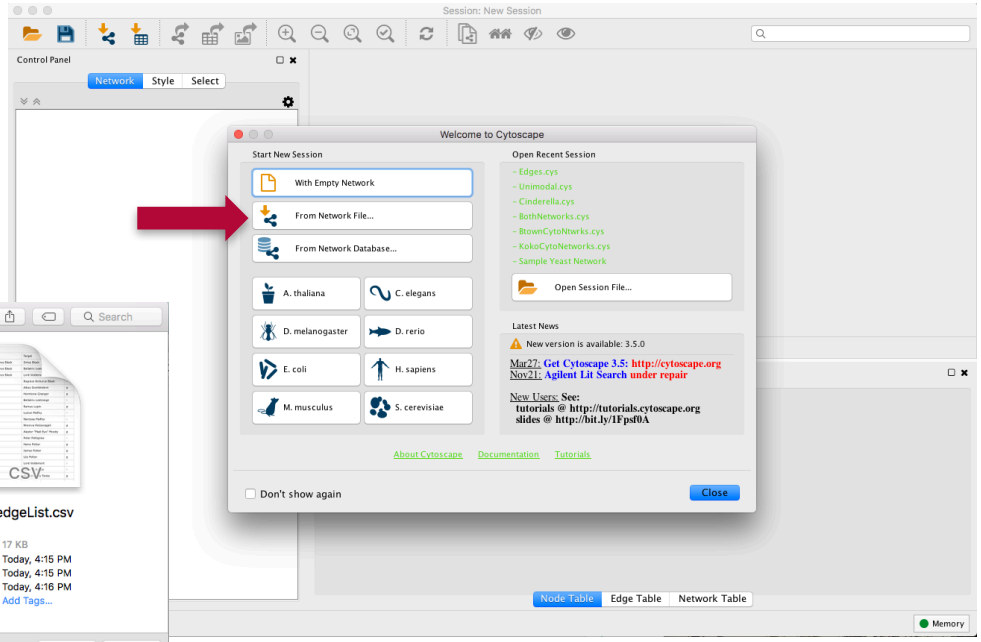
Save as HPedgeList.csv and HPnodeList.csv Respectively

NOTE: Better to download as .csv



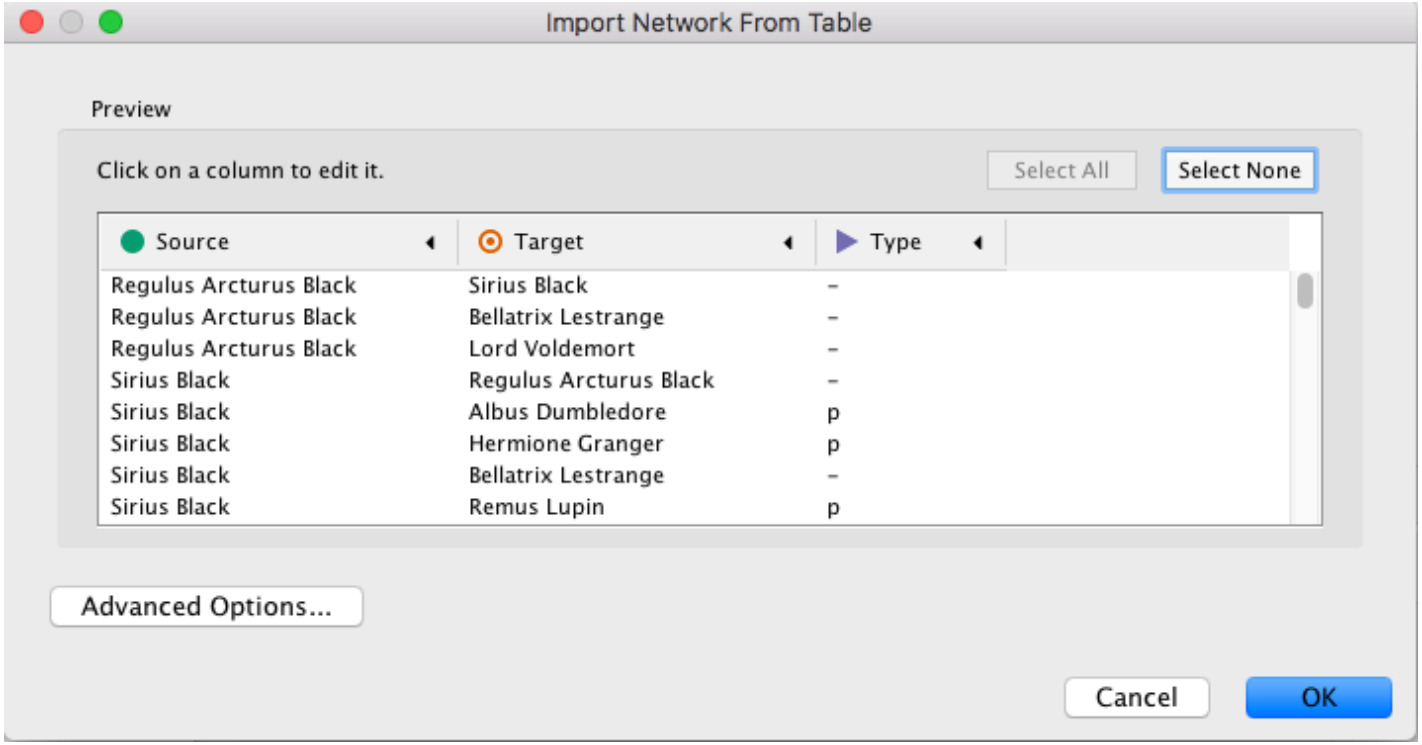
Upload .csv Edge List

- Open Cytoscape
  - Click "From Network File…" and choose your HPedgeList.csv file from the directory and/or folder you saved it in



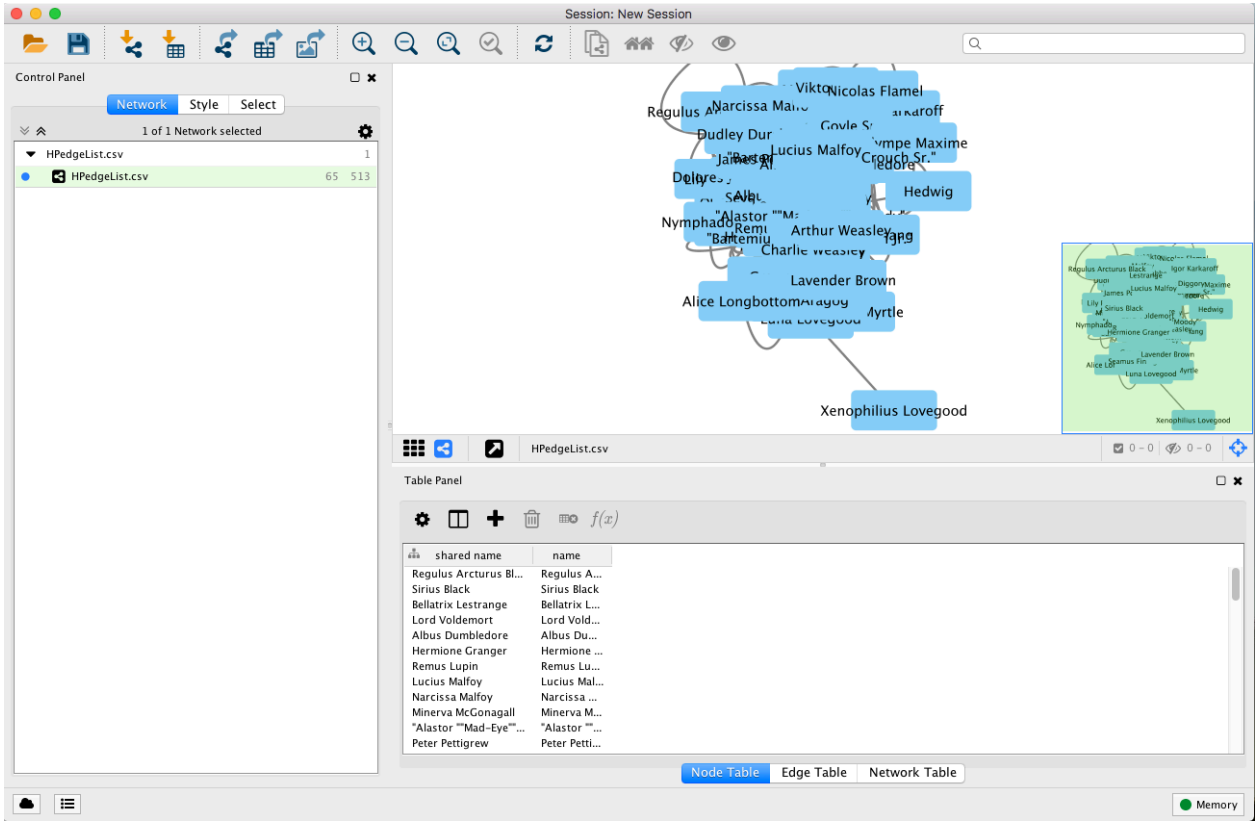
Source and Target

- Make sure the “source” column and “target” column are designated as such then click “ok”



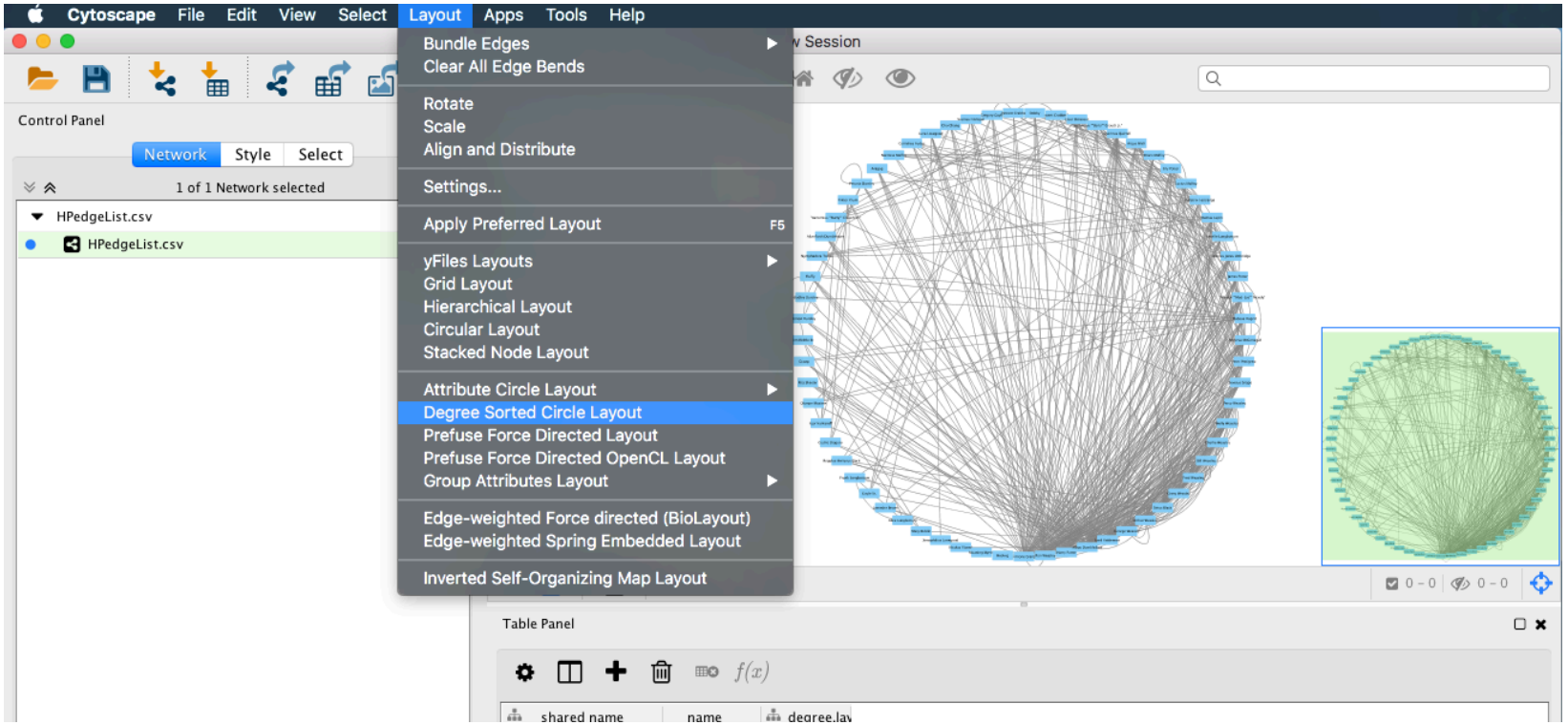
Hot Mess

- You should now have what appears to be a jumbled mess of spaghetti and meatballs.



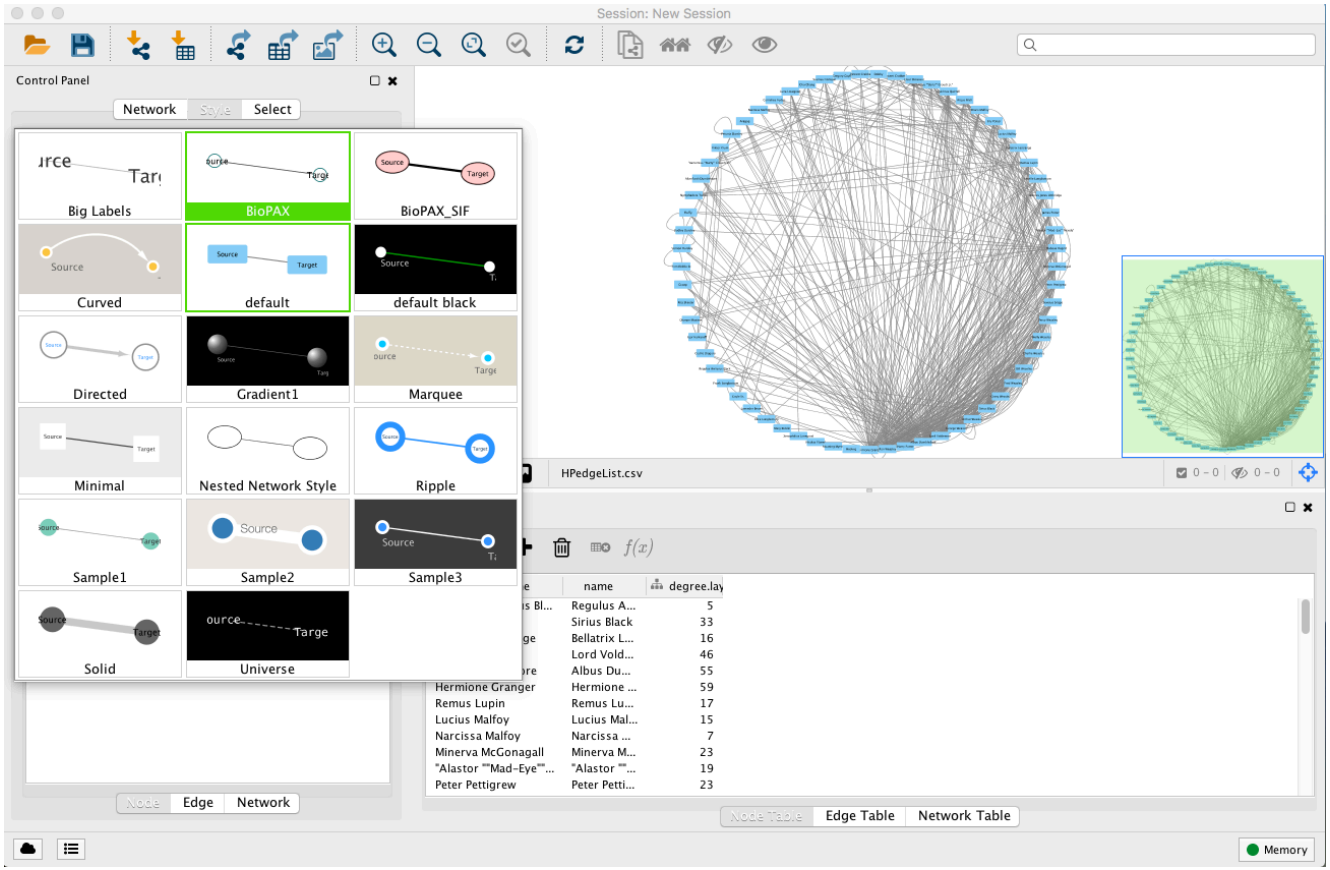
Layout

- Go to Layout>Degree Sorted Circle Layout (or one that you prefer)



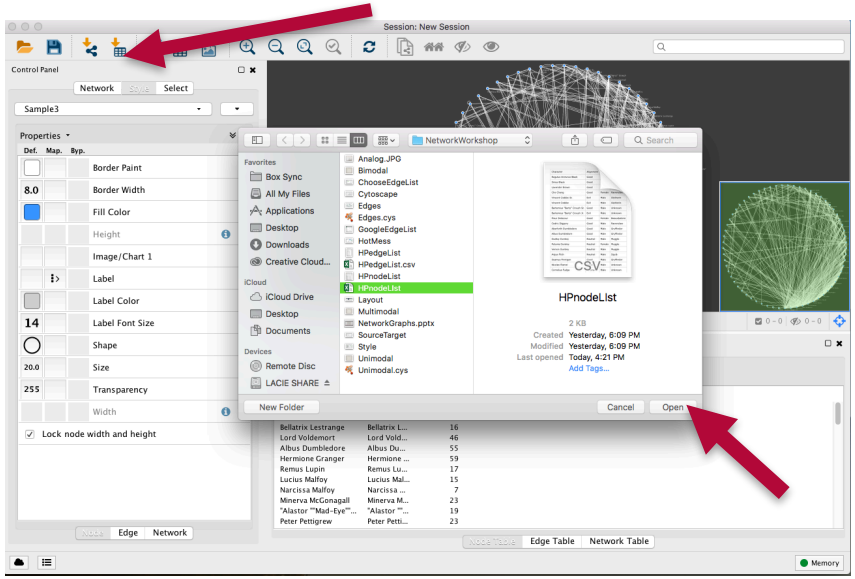
Style

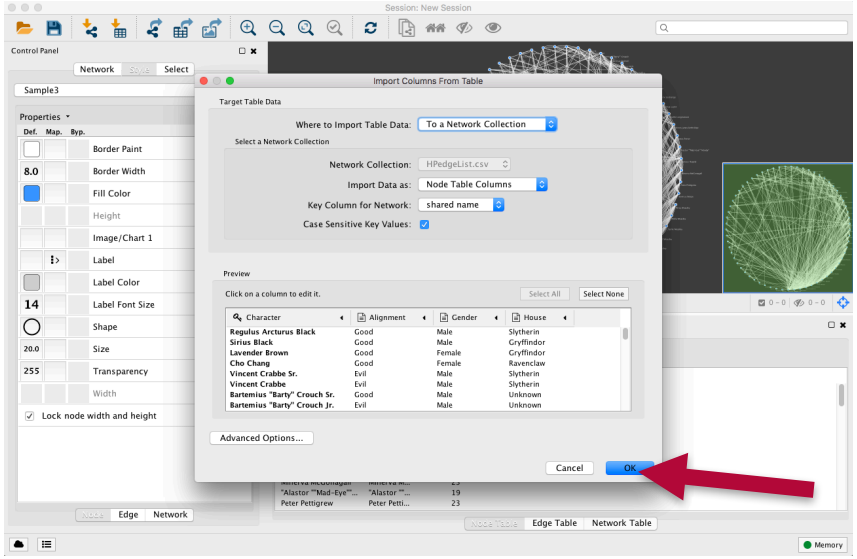
- This is "Sample 3" or feel free to choose one of your own



Upload Node List

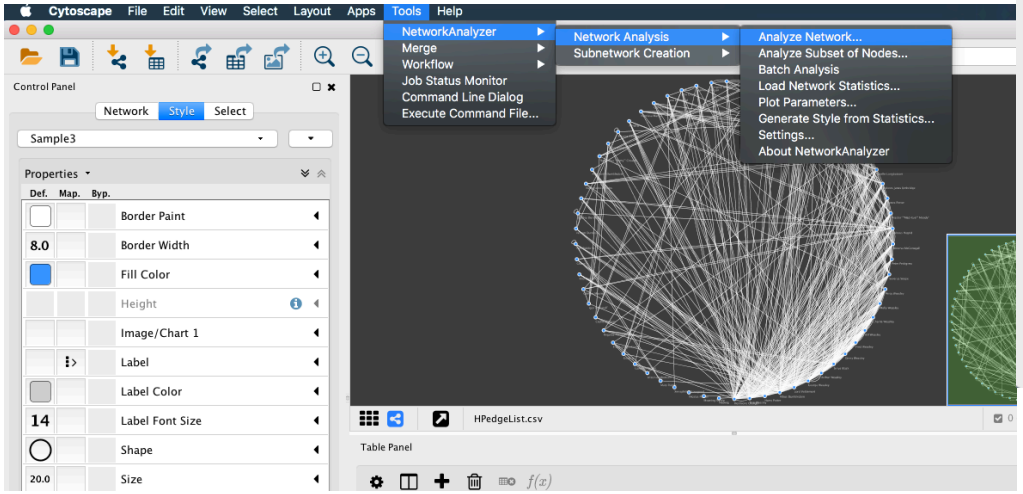
- Click on the “Import Table From File” icon, open file, then click “ok”

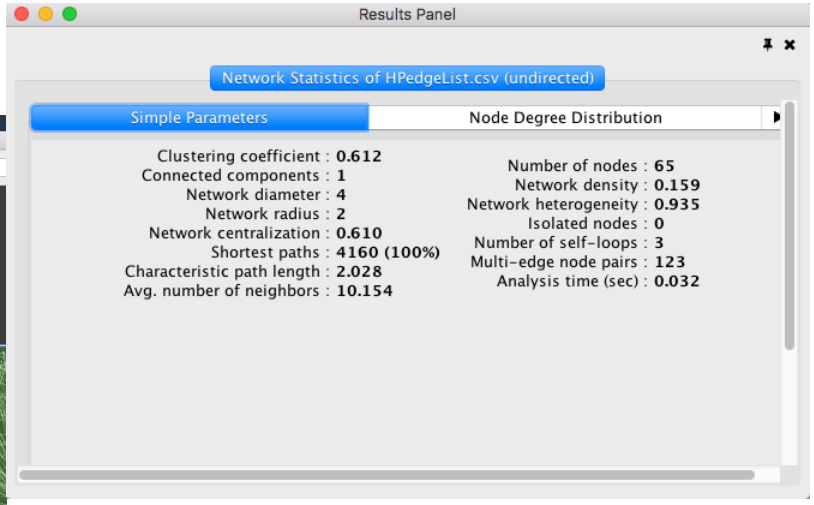




Analyze Network

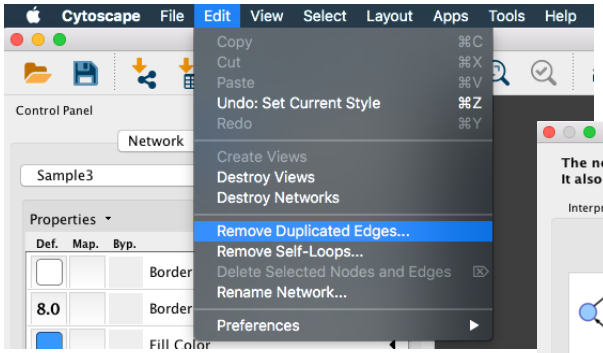
- Go to Tools>NetworAnalyzer>Network Analysis>Analyze Network
- The "Results Panel" should appear. Look at some of the terms and numbers.



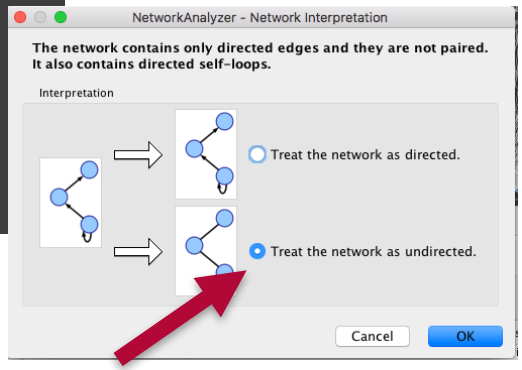


Remove Duplicates

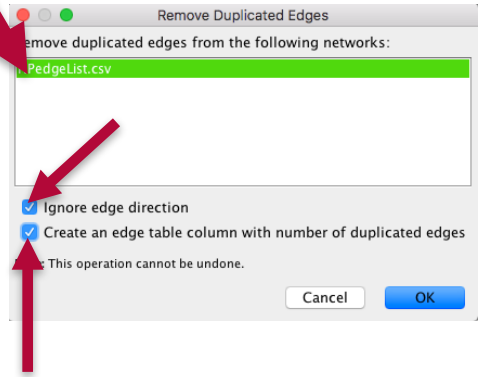
Go to Edit>Remove Duplicated Edges



Choose “Treat the network as undirected”

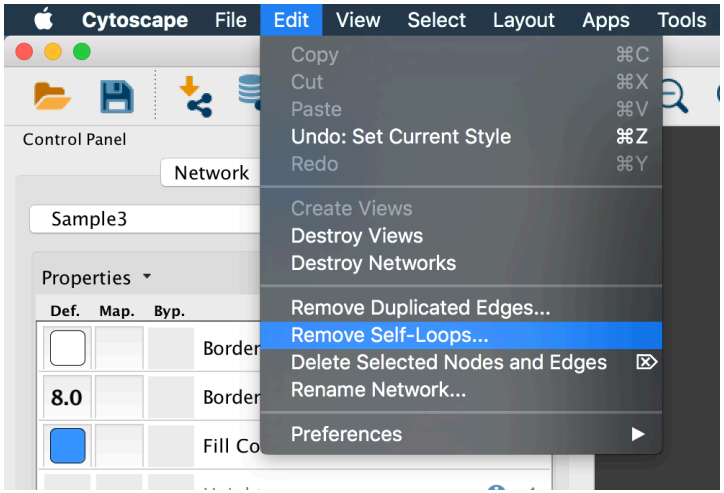


Make sure the file is highlighted and both boxes are checked

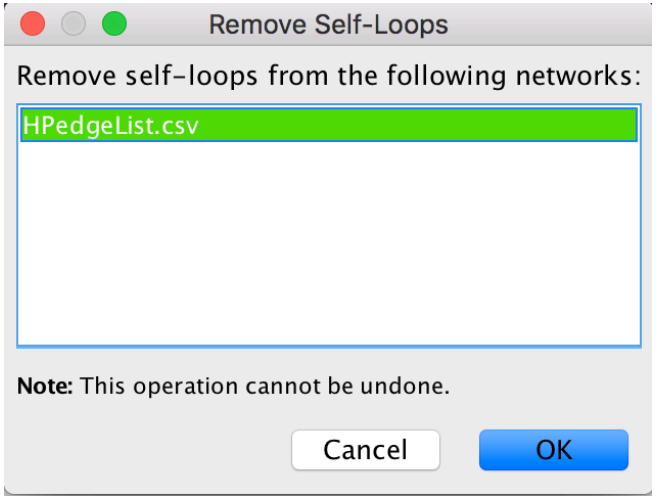


Remove Self-Loops

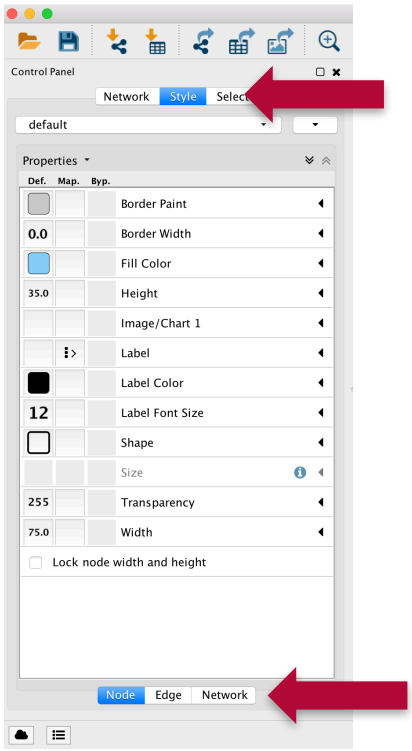
Go to Edit>Remove Self-Loops



Make sure file is highlighted
Then click "ok"



Let's Map



- To map the nodes and edges make sure you have "Style" selected on the top of the left panel
- Make sure you have the proper entity selected on the bottom as well. Either "Node" or "Edge", depending on which one you are interested in mapping
- Now let's see what this network can reveal…

NOTE: Don't be afraid to play around with the software or even break it.

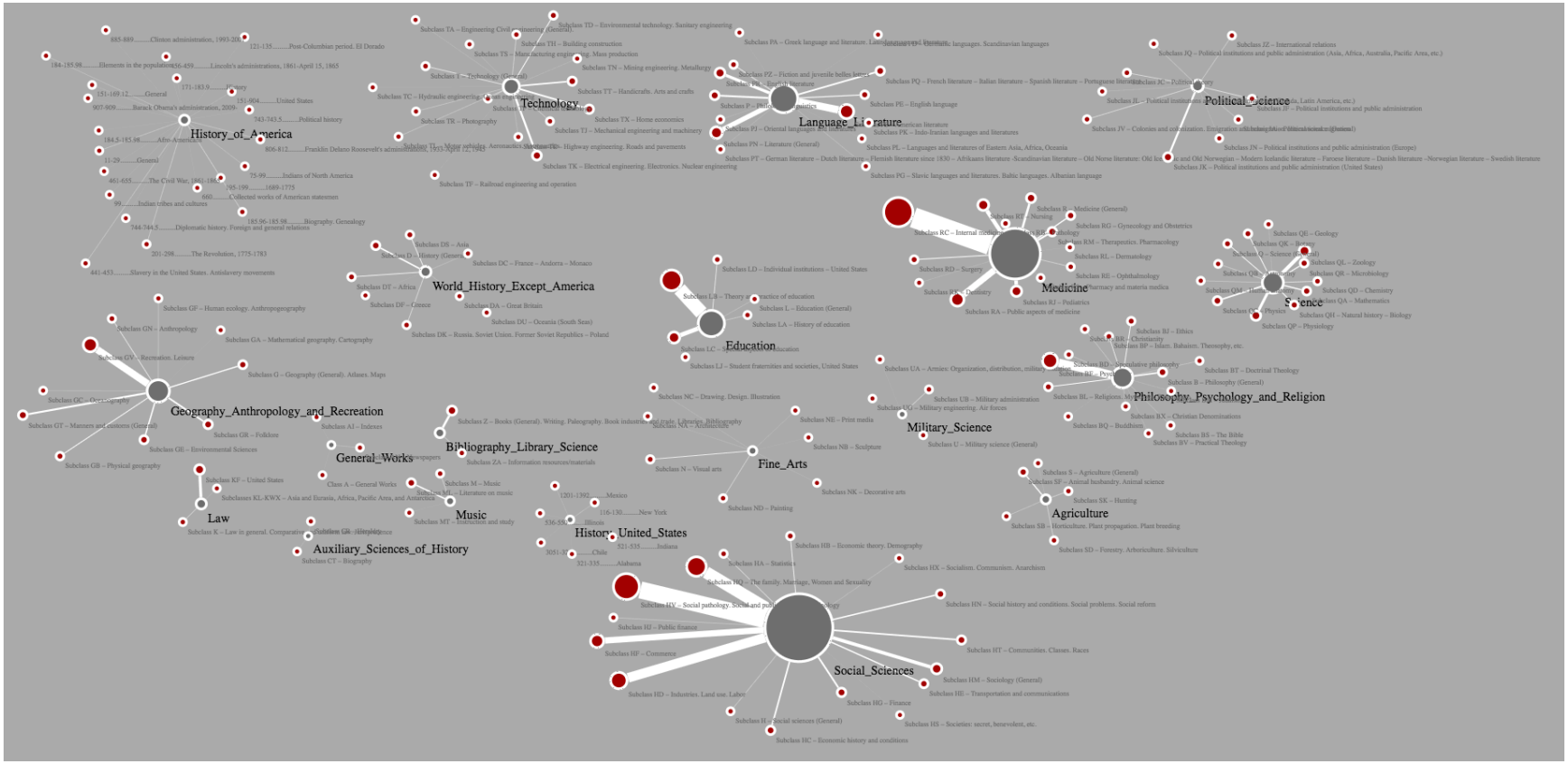
Important Things to Note

- Other visualizations may better reflect or help display your research:
  - May not accurately reflect your research (may be misleading)
  - May not be the best way to visualize your findings (not aid in understanding)

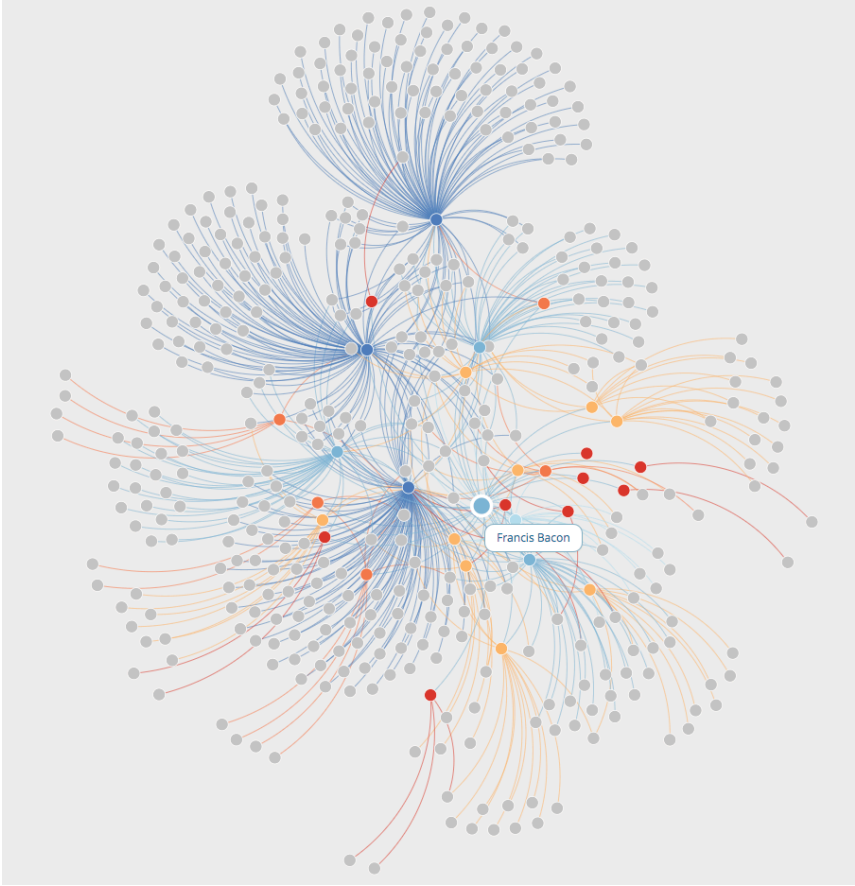
- Be aware of what the software you are using does:
  - May be made for other purposes (may need to alter steps)
  - May work better with other types of data (Geospatial networks)
  - May require more tech skills than you have or have time to acquire (since you have us, and many others that can help with this, it is not as big of a problem here at IU)

- Creating a network is not the end:
  - Keep careful track of every step and decision in making your network
  - Analyze the network carefully
  - Don't be afraid to start over (will almost always need to at least once, especially if you have a large network)

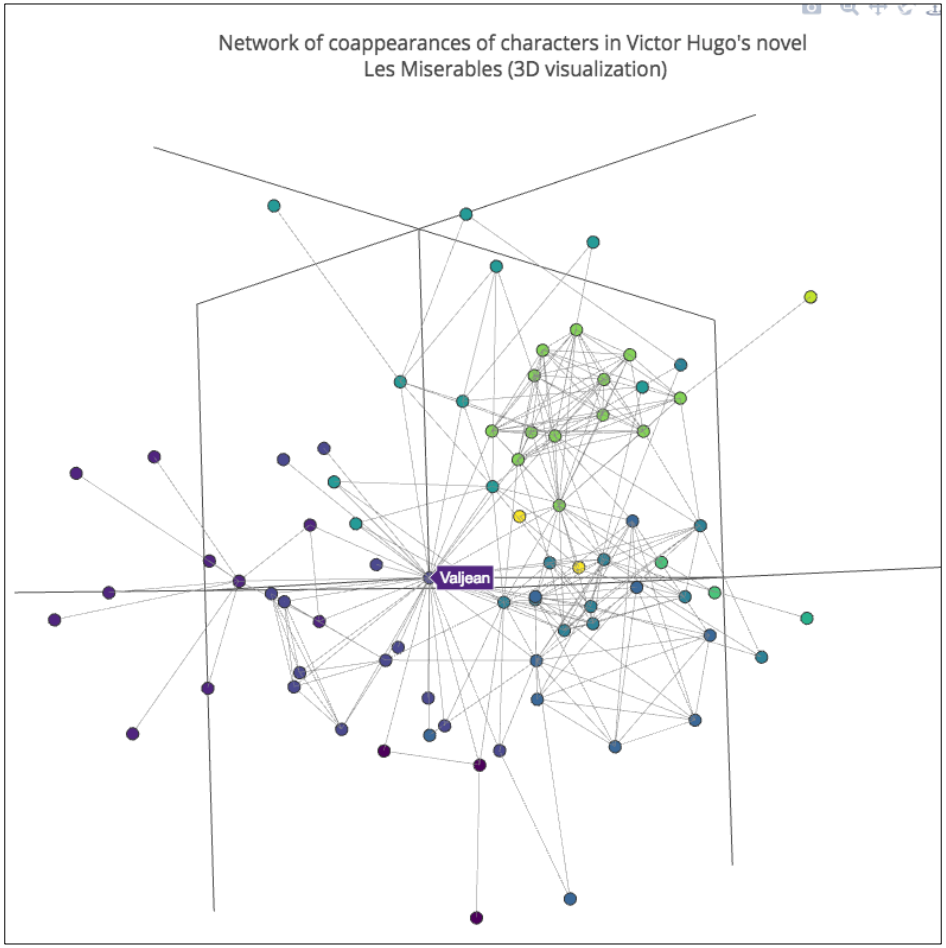
Sample Network:



Sample Network: Six Degrees of Francis Bacon



Sample Network: iGraph and *Les Miserables*



More Information

The CyberDH team can answer Network Graph questions. Get in touch!

Contact:
David Kloster
Cyberinfrastructure for Digital Humanities Coordinator
cyberdh@iu.edu

Cyber DH Blog